Available online at www.sciencedirect.com

SCIENCE DIRECT

Virology 348 (2006) 437–448

VIROLOGY

www.elsevier.com/locate/yviro

Minimization of genetic distances by the consensus, ancestral, and center-of-tree (COT) sequences for HIV-1 variants within an infected individual and the design of reagents to test immune reactivity

Girish S. Kesturu [a], Bonnie A. Colleton [d], Yi Liu [b], Laura Heath [b], Obaid Shakil Shaikh [a,c], Charles R. Rinaldo Jr. [d], Raj Shankarappa [a,d,*]

[a] *Thomas E. Starzl Transplantation Institute, Department of Surgery, School of Medicine, University of Pittsburgh, E1540 BST, 200 Lothrop Street, Pittsburgh, PA 15261, USA*

[b] *Department of Microbiology, University of Washington, Seattle, WA 98195, USA*

[c] *Division of Gastroenterology, Department of Medicine, School of Medicine, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA 15261, USA*

[d] *Department of Infectious Diseases and Microbiology, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA 15261, USA*

Received 6 September 2005; returned to author for revision 26 October 2005; accepted 30 November 2005
Available online 20 March 2006

Abstract

Eliciting maximal immune responses to highly divergent viruses is a challenge and a focus in AIDS vaccine development. Another challenge is to identify the immune correlates of protective immunity. Recent AIDS vaccine design approaches attempt to use reconstructed centralized viral sequences that minimize genetic differences to circulating viruses. Using these approaches, we derive and analyze consensus (CON), ancestral (ANC), and center-of-tree (COT) sequences to represent intra-individual HIV-1 *env* variants encoding a range of diversities and phylogenetic structures. Each reconstructed sequence significantly minimized genetic distances to extant sequences throughout the first 5 years of infection of an individual. Interestingly, ANC sequences diverged and were not significantly better than extant sequences in minimizing genetic distances at later stages of infection and disease, likely due to the development of a substantially asymmetric phylogeny. COT or CON sequences derived from autologous virus samplings may be useful for increasing the sensitivity of assessments of immune reactivity against HIV.

Keywords: HIV-1; HCV; Quasispecies; Genetic variation; Immune response; Cytotoxic T lymphocytes; Vaccine; Antibodies; Immunotherapy

Introduction

Viruses such as HIV-1, HCV, and others exist within an infected individual as a complex mixture of closely related but genetically distinct variants or the *quasispecies* (Eigen, 1993). Highly error-prone replication, unique to each of these viruses, leads to the generation and accumulation of such genetic variants. In HIV-1, the error rate of reverse transcriptase has been well characterized using various experimental systems (Mansky and Temin, 1995; Pathak and Temin, 1990a, 1990b; Roberts et al., 1988). In vivo, we have shown that HIV-1 consistently accumulates mutations at a rate of about 1% per year in its *env* (Shankarappa et al., 1999). A combination of such high rates of mutation and the turnover of billions of virus particles in a day (Ho et al., 1995; Wei et al., 1995) illustrates the enormity of the onslaught of virus mutations within each individual.

Assessing the impact of these mutations on viral biology and on viral interactions with the host helps to understand pathogenesis and to develop interventions. However, because of the huge number of mutations, studying the impact of variation requires new and practical approaches to reduce genetic complexity by deriving sequences that best represent the pool of variants. Since a large number of mutations encoded by these viruses are unique to each infected individual (Learn and Mullins, 2003), distinguishing the background mutations from

* Corresponding author. Thomas E. Starzl Transplantation Institute, Department of Surgery, E1540 BST, 200 Lothrop Street, Pittsburgh, PA 15261, USA. Fax: +1 412 624 6666.
E-mail address: RajShankarappa@gmail.com (R. Shankarappa).

0042-6822/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.virol.2005.11.055

the biologically relevant changes also helps in the design of an optimal sequence that can be used for immune reactivity and therapeutic studies. Here, we propose the design of these sequences using approaches currently being explored in the development of AIDS vaccine and evaluate their relationship to individual variants.

Evolution of HIV-1 at the inter- and intra-individual levels share a number of similarities. At the inter-individual level, a star-shaped phylogeny documented for HIV-1 represents the accumulation of mutations along multiple lineages on an ancestral backbone (Anderson et al., 2001; Myers et al., 1993). At the intra-individual level, clonal R5 HIV-1 establishes the infection (van 't Wout et al., 1994; Wolfs et al., 1992; Wolinsky et al., 1992; Zhang et al., 1993; Zhu et al., 1993, 1996) and accumulates mutations over time (Shankarappa et al., 1999; Wolinsky et al., 1996). Even when multiple variants initiate infection (Cornelissen et al., 1995; Learn et al., 2002; Poss et al., 1998), selective pressures appear to lead to a homogenization following the acute infection (Learn et al., 2002; Zhang et al., 1993). Therefore, accumulation of mutations within an individual immediately following the acute infection is qualitatively similar to an inter-individual star topology. However, over the long-term infection, mutations accumulate over successive generations of continually evolving populations and leads to a "ladder" shaped phylogenetic tree with temporally spaced variants (Mullins et al., 2004; Shankarappa et al., 1999; Wolinsky et al., 1996). Regardless, from the perspective of a single time point, star-phylogeny continues to best explain the phylogenetic relationship of variants at that time point because each variant generally represents a different set of mutations over an ancestral backbone sequence present at the earlier time point. In spite of these similarities, inter- and intra-individual viral evolution differs in a number of other important and potentially confounding ways. For example, the viral latency and compartmentalization are intra-individual phenomena that have not been defined at the inter-individual population level, but their equivalents are expected to exist. The level of inter-individual viral diversity is often much higher than intra-individual diversity.

The central role for immune system in the control of HIV-1 replication is incontrovertible. The hypothesis that genetic heterogeneity in HIV-1 is driven by a dynamic and ongoing recognition and escape of virus variants by the host immune system is consistent with the documented differences in the specificity of immune responses during acute and chronic stages of the disease (Goulder et al., 2001) and the appearance of differing escape mutations during early (Allen et al., 2000; O'Connor et al., 2002) and later (Kelleher et al., 2001) stages of infection. However, a number of other observations raise the possibility that currently used methods fail to provide a full and accurate description of anti-HIV-1 $CD8^+$ T cell responses. First, the breadth of viral mutations within an individual (Shankarappa et al., 1999; Wolinsky et al., 1996) cannot be explained by the escape at one or a few immunodominant epitopes. Second, mechanisms that lead to a large number of positively selected amino acid sites (Ross and Rodrigo, 2002; Yamaguchi-Kabata and Gojobori, 2000; Zanotto et al., 1999) are not understood. Finally, the use of autologous viral sequences (Altfeld et al., 2003), *toggled* peptides encoding variants (Korber et al., 2005), and the enhanced antigen presentation using dendritic cells (Huang et al., 2003; Zhao et al., 2002) appears to identify stronger and new $CD8^+$ T cell responses that would otherwise be missed. These arguments are also strengthened by a recent report that links immune recognition with nearly two-thirds of all persisting mutations observed in individuals (Allen et al., 2005).

In the development of AIDS vaccine immunogens, issues concerning viral diversity are being tackled by reconstructing centralized immunogen sequences that minimize the genetic distance to circulating viruses (Gao et al., 2004; Gaschen et al., 2002; Learn and Mullins, 2000; Mullins et al., 2004; Nickle et al., 2003a; Novitsky et al., 2002). By being genetically closer to each of the circulating variants, such an immunogen is expected to elicit the maximum immune response and thus hoped to be protective. Two such computer generated *env* sequences have been recently shown to confer functional biologic properties to the virus and elicit immune responses that can recognize and neutralize some strains of HIV-1 (Doria-Rose et al., 2005; Gao et al., 2005). We propose that both the eliciting and testing of anti-HIV-1 immune responses are two sides of the same issue and requires the use of similar approaches. Unlike the vaccine context, each infected individual harbors genetically distinct virus and a population of variants whose composition changes over time. Therefore, it is necessary to test the extent to which the methods aimed at minimizing the genetic distances apply to intra-individual variants exhibiting differing levels of diversity and phylogenetic relationships observed over the course of infection. Here, we derive and characterize the ability of three centralized representative sequences, consensus, ancestral, and COT sequences, to minimize the genetic distances for HIV-1 *env* sampled from 13 consecutive time points over an 8-year period in an individual. These sequences represented a wide spectrum of changes typically observed in an individual. We show significant differences in the phylogenetic relationship among variants present at different times in the infection and document its effects on the minimization of genetic distance. We also propose the use of phylogenetic properties to test hypotheses related to immune escape and the use of genetic analyses for natural selection in designing reagents to test immune reactivity.

Results

Impact of diversity on the minimization of genetic distances

The goal of sequence reconstruction methods is to design a sequence that minimizes the genetic distance to a set of variant sequences. Therefore, we first sought to test whether the extent of this minimization in the genetic distance is impacted by changes in levels of diversity. Sequences used in this study represented a broad spectrum of intra-individual nucleic acid diversity, ranging from 0.2% (0.4% for amino acid sequences) in the first post-infection time point to 6.2% (11.4% for amino acid sequences) at about 6 years following infection. Fig. 1 illustrates the average values for pairwise nucleic and amino

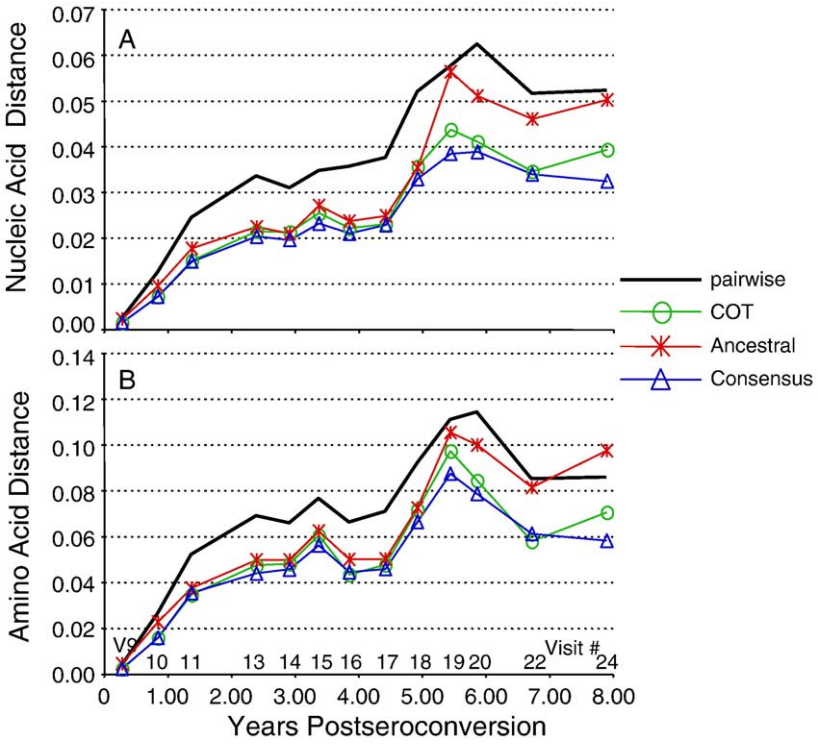

Fig. 1. Minimization of genetic distances by the consensus, ancestral, and COT sequences at different time points over the course of infection. Genetic distances for the nucleic acid (panel A) and amino acid (panel B) sequences were derived using the Kimura 2-parameter model for evolution and as *p*-distances, respectively. Genetic distances corresponding to the average for the pairwise comparisons between individual variants are depicted as thick black lines. The average genetic distances between each representative sequence and the individual variants at the corresponding time points are plotted as lines with symbols as identified in the figure.

acid diversity over the course of infection and its reduction by the corresponding consensus, ancestral, and COT sequences. All three sequences reduced the genetic distances over the range of viral diversity observed for the first 5 years. Subsequently, ancestral sequences failed to minimize distances and were unlike that for COT and consensus sequences. At these visits, COT and consensus sequences also showed a departure from the close concordance seen at the earlier visits.

The minimization in genetic distances by each of the representative sequences was quantified to allow a better comparison across time points and is illustrated in Fig. 2. Overall, the COT sequences led to an average minimization of 28 ± 8% (mean ± SD) for amino acid sequences and 33 ± 5% for nucleotide sequences. Analogous values for the consensus sequences were 31 ± 5% for amino acid and 37 ± 3% for the nucleic acid sequences. In contrast, the extent of minimization of genetic distances was much reduced for the ancestral sequences (13 ± 16% for amino acid and 21 ± 13% for the nucleic acid sequences). As evident in Fig. 1, the discrepancy in the minimization of genetic distance between the ancestral and other sequences was strongest for the last four visits. Consistent with this discrepancy, the number of amino acid differences between the COT and ancestral sequences also increased significantly (five or less amino acid differences for V09–V17 vs. 13–18 amino acid differences for V18–V24). These findings allowed us to conclude that the consensus and COT sequences minimize genetic distances to similar extents over a range of levels of diversity observed during the course of infection. In contrast, the ability of ancestral sequences to minimize genetic distances can be different for sequences sampled from different time points in the same individual.

Phylogeny of representative sequences

We next sought to evaluate the phylogeny of variants sampled within each time point along with their representative sequences as well as to test for a phylogenetic explanation for the inability of ancestral sequences to minimize genetic distances in the subset of later time points. In addition, since the consensus, ancestral, and COT sequences were derived separately using the subsets of sequences, we needed to test their phylogenetic behavior when all the sequences were included in the analysis. Fig. 3 illustrates the phylogenetic relationship among sequences sampled from each time point along with the three centralized representative sequences for each corresponding time point. To provide a uniform frame of reference across different comparisons and since ancestral sequence is the explicitly designed parental sequence for each group, each phylogram was rooted to the ancestral sequence. The phylogram format, as contrasted with the radial format, also helped us to better separate and illustrate the lineages.

The phylograms in Fig. 3 provide an overall snapshot of the changes in phylogenetic relationship among viruses sampled at each visit. The tree topology was generally similar when comparing adjoining time points, and the dissimilarity between phylograms increased with an increase in the length of time

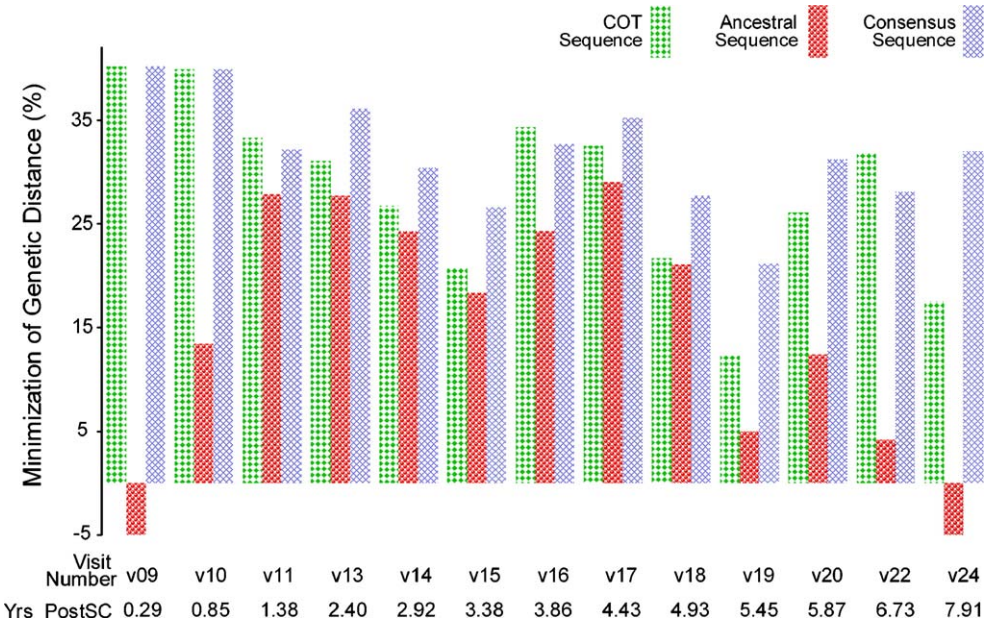

Fig. 2. Differences in the minimization of amino acid distances by the consensus, ancestral, and COT sequence. Data illustrated in panel B of Fig. 1 were converted to percentage differences and plotted. COT and consensus sequences minimized genetic distances to similar extents, whereas ancestral sequences varied in their ability to minimize the distances. This dichotomy was pronounced for sequences sampled from the last four visits.

separating them. Even though no two phylograms were completely identical, the phylograms could be classified into two broad groups. In the first group (V09–V18), multiple branches of similar branch lengths separated individual taxa or major clusters, each projecting from a vertical backbone and leading to an overall balanced shape. In contrast, a wide range of branch lengths separated taxa in the second group (V19–V24). According to the terminology previously described (Mullins et al., 2004), the first group is analogous to a combination of balanced, star, and lopsided tree shapes, whereas the second group represents the ladder shape.

Consistent with the proposed models for phylogenetic clustering of each representative sequence (Mullins et al., 2004; Nickle et al., 2003a), consensus and COT sequences clustered towards the middle of the tree, whereas the ancestral sequences were off-centered in unrooted trees (not shown; Fig. 3 shows trees rooted to the ancestral sequences). During the first 5 years of infection, small branch lengths separating the three

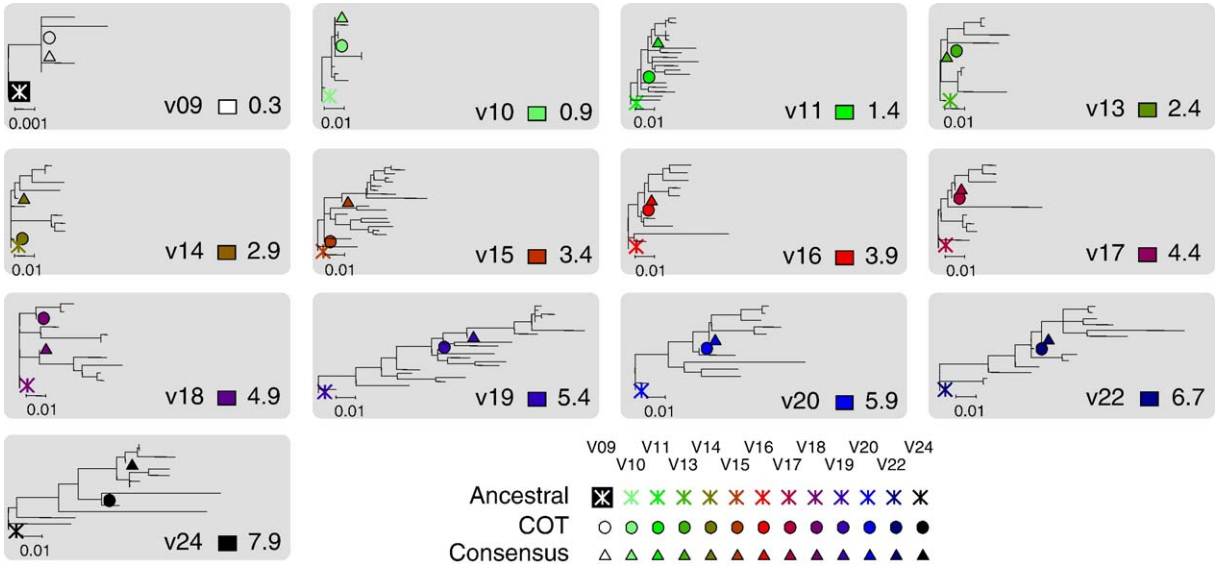

Fig. 3. Phylogenetic relationship among sequences sampled at each time point and the corresponding consensus, ancestral, and COT sequences reconstructed to represent the sequences sampled at each time point. Neighbor-joining phylograms were generated using maximum likelihood distances estimated for each data set. Each phylogram was rooted to the ancestral sequences to provide a uniform frame of reference across different phylogenies. Due to low genetic diversity, V09 sequences were drawn to a different scale. Numbers next to the symbol within each panel represent the years following seroconverson. The asymmetric phylogeny in the last four visits appears to strongly influence the branch lengths that separate the ancestral from other representative sequences.

representative sequences reflected a small genetic difference between them (only horizontal lengths are used to measure genetic differences). These time points were also characterized by low diversity and a relatively balanced tree topology. In contrast, high diversity and asymmetric topology, as observed for V19–24, led to long branch lengths separating the ancestral sequence from the other two.

A temporal ordering of the variants on a phylogenetic tree, as we had previously described (Shankarappa et al., 1999), reflects a dynamic turnover in the virus population. New mutations are added in an iterative manner to the continually evolving backbone sequence. With each sampled time point representing a portion of this continuum, sequences designed to represent one such portion would be best defined by a position at the center of its tree space. Because of the temporal features inherent to these sequences, the reconstructed centers of the tree space could be shifted back with respect to time. Therefore, representative sequences from one time point could even be more closely related to variants from a prior time point, especially for an ancestral sequence given the explicit design criteria. Fig. 4 illustrates a test of this reasoning by evaluating the difference in genetic distances between the representative sequences and the individual variants from its contemporary vs. the prior time points. A preponderance of positive values for the current dataset suggests that, on average, representative sequences are closer to sequences from the contemporary time points. Such closer relationship could be due to sampling frequency and/or the rate of genetic change. However, the ancestral sequences from a few time points did appear to be more closely related to sequences from the prior time points. Such relationship for the early ancestral sequence could be related to the homogenization of sequences following the acute infection (Learn et al., 2002) and/or an evolution towards the ancestral state (Herbeck et al., 2006). Such a relationship for the later visits (V19 and V24) could be the result of phylogenetic asymmetry, due to contributions from different compartments (Nickle et al., 2003c).

We next evaluated the phylogenetic relationship between all the 189 sequences used in this study by including the consensus, ancestral, and COT sequences derived for each time point along with the 150 individual variants in a single phylogram (Fig. 5). This analysis was undertaken to examine how each representative sequence for different time points phylogenetically relate to individual variants. We had previously documented the strong temporal structure among the individual variants (Jensen et al., 2003; Shankarappa et al., 1999), and the inclusion of additional sequences did not appear to perturb the temporal relationship. In addition, representative sequences derived from each time point also clustered around variants from the corresponding time points. Therefore, the derivation of representative sequences did not induce any artifactual changes causing the sequences to cluster in an aberrant manner. Similarly, each representative sequence encoded genetic features that were adequate and specific to lead them to cluster with the sequences they were derived from. Additional validation of this reasoning was found when each set of representative sequences showed a rough clustering according to the time of their sampling (in the inset of Fig. 5). A perturbation of this temporal order seen for sequences from the later time points is consistent with the asymmetric phylogeny for sequences sampled at these time points.

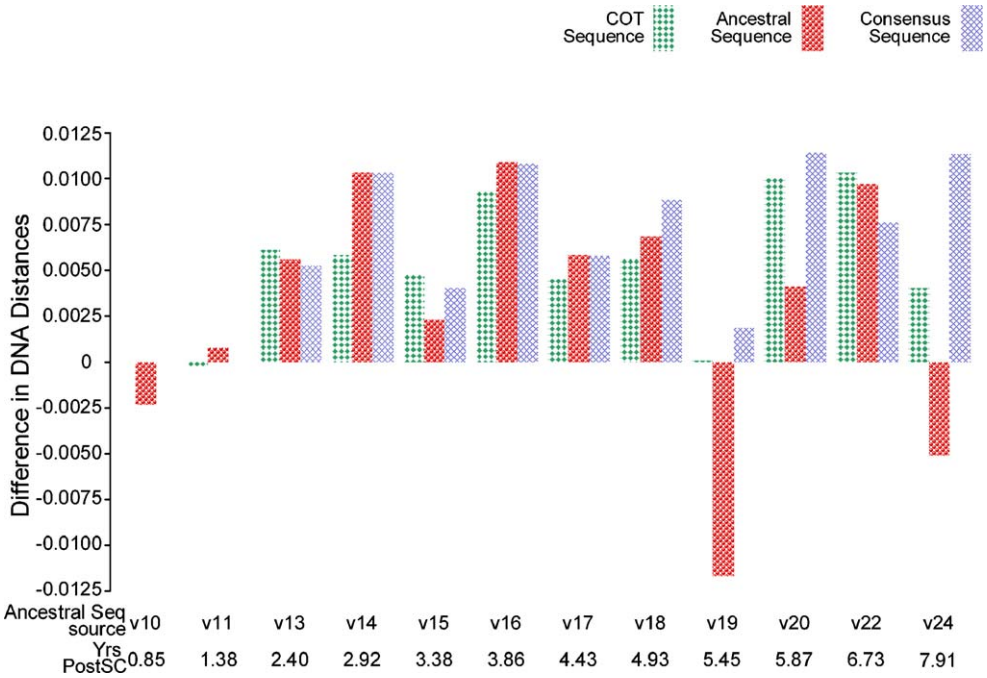

Fig. 4. Comparison of variants from contemporary vs. prior time point for their genetic similarity to the contemporary consensus, ancestral, and COT sequences. Genetic distances between representative sequences for each visit and the individual sequences for the contemporary and the prior visits were derived, and the differences were plotted. Positive values indicate that representative sequences are genetically closer to the contemporary sequences they were derived from. Negative values for visits 10, 19, and 24 indicate that these ancestral sequences were on average, genetically closer to sequences sampled from the prior visit.

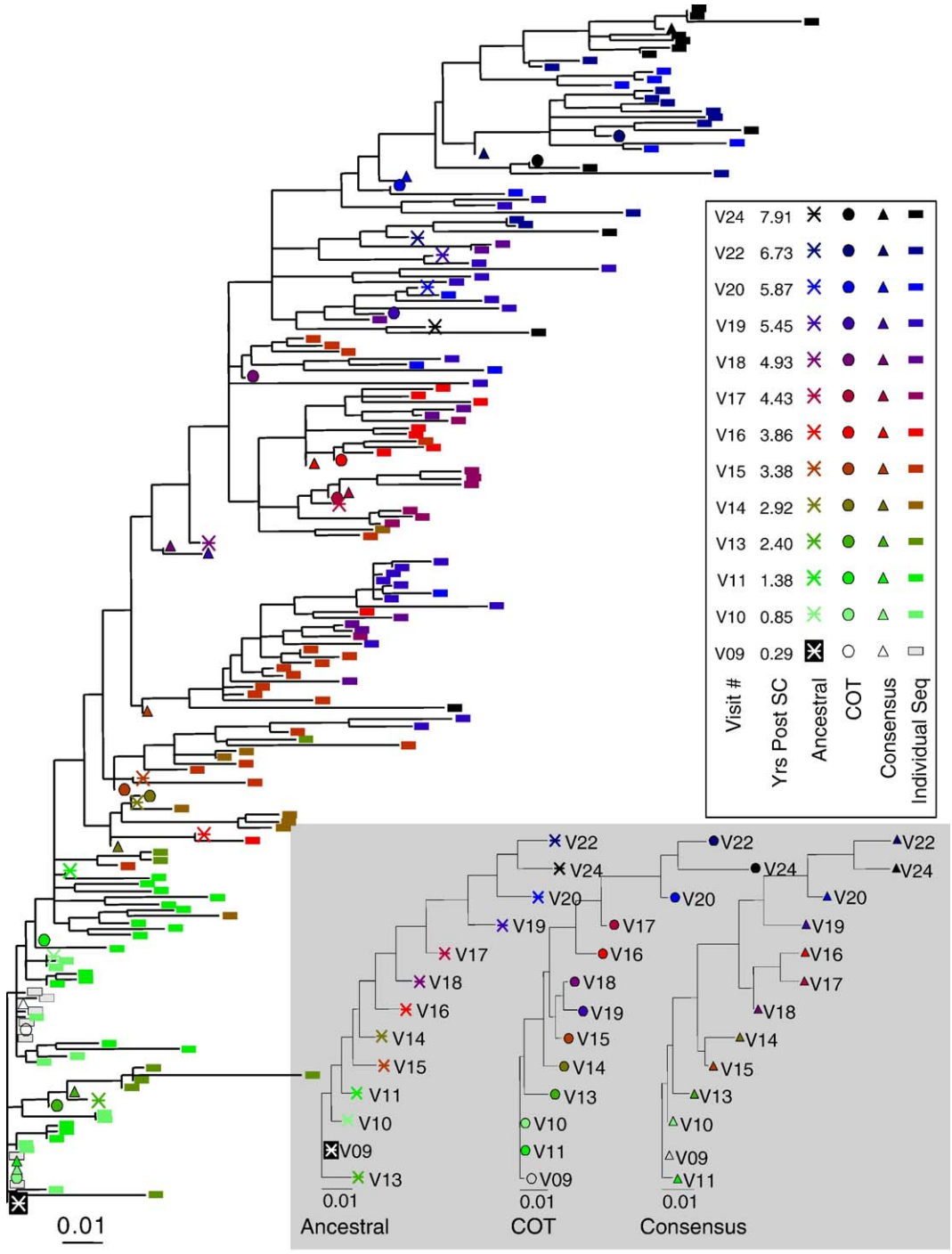

Fig. 5. Phylogenetic relationship among all the individual variants sampled over the 8-year/13-visit follow-up, along with the 39 representative sequences derived for the pools of sequences sampled at each time point. Sequences were aligned to match codons, and the neighbor-joining phylogram was derived using model parameters deduced for the entire dataset. Each of the representative sequences from different time points generally clustered closer to individual sequences from the contemporaneous time point and were consistent with the overall temporal relationships observed for individual variants. The exceptions appear to be the ancestral sequences corresponding to the last four time points that clustered close to each other. The inset illustrates the phylograms derived separately for the consensus, ancestral, and COT sequences representing each time point.

Selecting sequences to test for immune escape

The centralized representative sequences derived and analyzed earlier are appropriate targets to test immune reactivity to autologous virus sequences. However, the cost and other experimental considerations, such as the availability of cryopreserved cells, may make it difficult to test immune reactivity to the entire panel of peptides corresponding to the autologous viral sequence. Therefore, limiting the testing of reactivity to regions corresponding to known cytotoxic T

lymphocyte (CTL) epitopes is one option to reduce the number of candidate peptides targeted for screening. However, this approach ignores the main reasons why autologous viral sequences are needed to test the full immune reactivity, i.e., the effect of high inter-individual differences on the recognition of new and unique CTL epitopes by different individuals. Therefore, it is highly desirable to use additional and potentially alternative methods to choose regions within the autologous sequences that may be of immunologic significance.

By definition, ancestral sequence traces the path to a point that represents the parental sequence for all the individual variants. Under certain conditions such as an iterated accumulation of mutations, an increasingly heterogeneous group of variants could lead to an ancestral sequence that is progressively more divergent from the variants. Therefore, the reduced minimization of genetic distances by the ancestral sequences, as compared to the consensus and COT sequences, is consistent with this reasoning. In fact, for certain pools of intra-individual virus variants, instead of minimizing the distances, ancestral sequences led to an increased average genetic distance, or the ancestral sequence from one time point was more closely related to variants from an earlier time point (Figs. 2 and 4). Therefore, while the ancestral sequence attempts to encode sequence elements present in the past, the COT and consensus sequences reflect features most common to the set of contemporary sequences. If the continual escape from immune recognition is an ongoing iterative process, then the ancestral sequence has a higher likelihood of encoding sequences present prior to immune escape, whereas the consensus and COT sequences encode changes that allowed the virus to escape immune recognition. Based on this reasoning, we hypothesize that the COT and consensus sequences are best suited to test the breadth of immune reactivity to contemporary viruses, and the ancestral sequences are most appropriate to test immune reactivity present prior to immune escape. It is possible that a large fraction of differences between the ancestral and COT sequences may in fact represent the actual mutations that allowed the virus to escape immune responses. A reduction in the number of candidate sequences made possible by the use of representative sequences makes this a practical and rational approach.

Testing CTL reactivity using the above strategy relies on the extent of differences between COT and ancestral sequences. In this study and until V18, we found one to five amino acid differences between ancestral and COT sequences, and this difference increased to 13–18 amino acids thereafter. When these differences are small, as for the sequences from earlier time points, sequences from multiple time points can be used to derive an ancestor that can take one farther back along the tree. On the other hand, if the differences are large, as in the instances of contributions from the reservoir/compartments, it may be necessary to identify such sequences (Nickle et al., 2003b) and generate subsets to represent the contemporary and archival variants. In general, the use of consensus or COT sequences appears to be most informative for early infection samples, whereas coupling the use of ancestral sequence with COT/consensus sequence will be more informative in identifying potential escape mutations.

We also propose a second approach to choose reagents to test for immune escape that is built around the argument that immune recognition drives natural selection at selected sites of the sequence. Therefore, identifying sites that show evidence for positive selection identify reasonable targets to test for immune recognition and escape. This argument is supported by the dominant presence of positive selection over the course of infection and its relationship to epitopes recognized by T lymphocytes (Liu et al., 2004; Nielsen and Yang, 1998; Ross and Rodrigo, 2002; Williamson, 2003; Zanotto et al., 1999). In Fig. 6, we illustrate the amino acid sites that show evidence for positive selection over the course of infection. We find multiple contiguous positively selected sites downstream of V3, within the V4 and V5 regions, as well as an overall increase in the number of such sites over the course of infection. Many of these sites also showed evidence for consistent positive selection over the course of infection. Positively selected sites were also observed within the two known CTL epitopes restricted by Pt 8's HLA (Korber et al., 2003). We propose that peptides derived from regions that exhibit positive selection are candidates to test for immune reactivity and are hypothesized to contain variants that allow the virus to escape immune recognition. Since our goal in this study was to identify and display all potential sites under selection, we used less stringent criteria. In another study, Ross and Rodrigo (2002) have reported a more rigorous assessment of the positive selection, and under varying selection criteria and values of ω (d_N/d_S), for sequences from this and other individuals in our previous study (Shankarappa et al., 1999).

The final selection of peptides to test immune reactivity can rely on a combination of the above approaches. The use of these approaches may also depend on the type and frequency of sampling sequences, level of variation, and the number of candidate motifs identified. For example, the availability of a large number of sequences sampled using rigorous criteria, such as those involving care to prevent resampling of virus templates, will allow an accurate reconstruction of the ancestral (Doria-Rose et al., 2005) and COT sequences as well as test for positive selection. If sequences from additional time points are available, positively selected sites can be tracked over time to test for fixation of mutations and to better define peptide candidates to screen for immune escape.

Discussion

Deriving a centralized reconstructed sequence to represent a population of variant sequences for use in vaccine is a novel approach and is being pursued aggressively in the development of an AIDS vaccine immunogen (Doria-Rose et al., 2005; Gao et al., 2004; Gaschen et al., 2002; Learn and Mullins, 2000; Mullins et al., 2004; Nickle et al., 2003a; Novitsky et al., 2002). The derivation of reconstructed sequences at the inter- and intra-individual levels has analogous conceptual bases but has not been reported for the intra-individual variants. Here, we show that early stages of chronic infection are characterized by phylogenetic symmetry and similar minimization of genetic

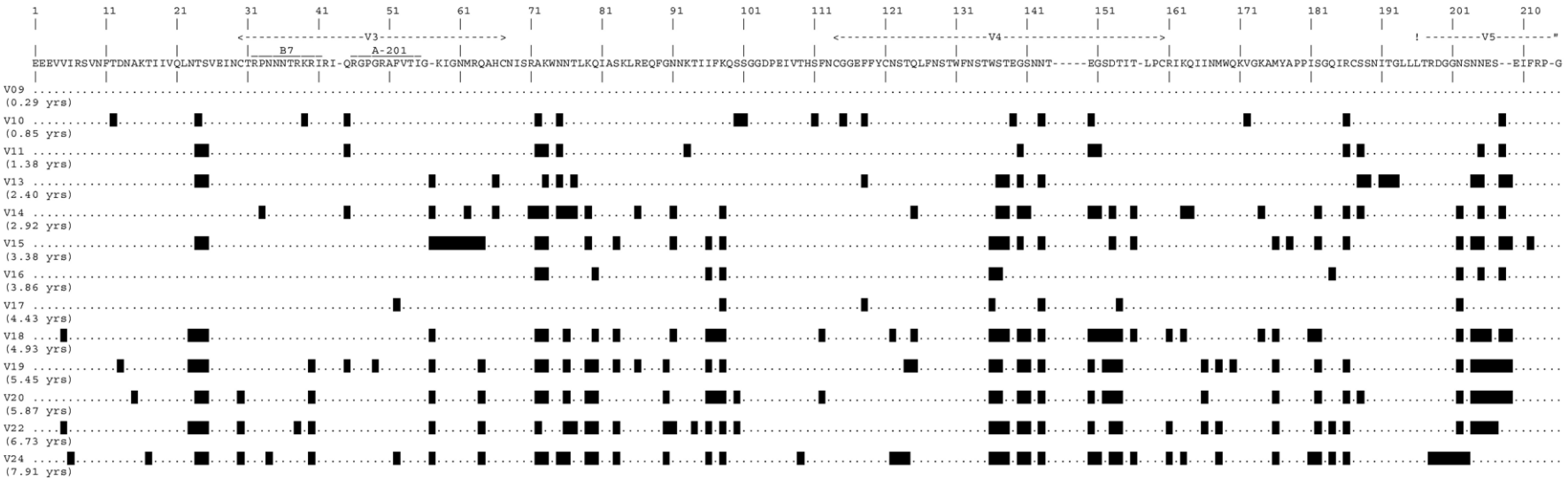

Fig. 6. Changes in amino acid sites exhibiting evidence for positive selection over the course of infection. The consensus amino acid sequence is shown at the top and the identity of the time point is shown on the left. Highlighting the position corresponding to the amino acid at each visit identifies sites that show evidence for positive selection. Also illustrated are the positions corresponding to the three variable regions and the two CTL epitopes restricted by HLA-A0201 and -B7, encoded by this individual.

distances by the consensus, ancestral, and COT sequences. On the other hand, variants from later stages of disease exhibit a more complex phylogeny, marked by asymmetry and potential contributions from the reservoirs and compartments. These changes can lead to significant differences in the minimization of genetic distances by the ancestral vs. other methods of reconstructing centralized representative sequences. The presence of these differences illustrates the changing landscape of viral evolution over the course of infection. These changes also underscore the importance of undertaking these analyses in choosing autologous sequences, as opposed to choosing a cloned sequence at random, for immune reactivity and immunotherapeutic studies.

We propose that eliciting and detecting maximal immune responses are two sides of the same issue, and hence, the approaches being used in one are applicable to the other. Just as high viral diversity has impacted the development of a vaccine, we contend that a full understanding of the correlates for protective immune reactivity has also been influenced by the differences between viral sequences used to test immune responses and those present in the infected individual. The 30% enhancement in the number of CTL epitopes identified using autologous vs. subtype consensus sequence (Altfeld et al., 2003) could be an underestimate for two reasons. Other genes such as Env that were not used in this study are much more divergent, and the use of enhanced antigen presentation by dendritic cells can elicit stronger and new immune reactivities. A more parsimonious explanation for the clustering of CTL epitopes to conserved regions (Yusim et al., 2002) is in the identification of most epitopes using reagents based on the prototype sequences. A CTL response that is somehow most pronounced to a clustered set of conserved regions is not consistent with a highly complex profile of T cell responses (Addo et al., 2003; Altfeld et al., 2003; Frahm et al., 2004; Kaufmann et al., 2004; Yu et al., 2002). The high levels of variability observed within a time point in an individual and the continual replacement of the population of variants by the new and continually evolving variants (Shankarappa et al., 1999; Wolinsky et al., 1996) are phenomena consistent with an active and dynamic recognition by the immune system. But this dynamism will not be apparent if the testing of immune reactivity fails to consider the sequence differences.

We document significant differences in viral diversity and phylogenetic relationship among sequences sampled at different time points within an individual. Therefore, the use of methods such as deriving a consensus sequence by bulk sequencing of the variant pool would not be appropriate when the underlying phylogenetic relationships differ. The choice of representative sequence must also consider factors that underlie and influence their derivation. Consensus sequence is derived using a simple concept and methodology that links the most frequent or a modal nucleotide or amino acid at each site. However, this derivation is strongly influenced by the sampling artifacts: inadequate sampling of sequences may lead to a biased consensus and the artifactual resampling of viral templates may lead to a wrong consensus sequence. In contrast, deriving COT sequence involves the use of models of evolution to reconstruct the sequences (Nickle et al., 2003a) and hence is less susceptible to sampling artifacts. On the other hand, ancestral sequences are influenced by a different set of parameters and are best illustrated by the effect of phylogenetic asymmetries (Mullins et al., 2004; Nickle et al., 2003a). As previously pointed out (Gao et al., 2003; Nickle et al., 2003a) and illustrated here for V11 through V17 sequences, all three sequences do not differ in the minimization of genetic distances in the presence of near-symmetric phylogenies. However, even though the differences between the three sequences have been proposed to be of little statistical significance at the population level and less relevant in the design of a vaccine (Gao et al., 2003), we point out that at the intra-individual level, these differences are highly relevant and must be considered.

In summary, we derive the consensus, ancestral, and COT viral sequences to represent heterogeneous sets of HIV-1 variants sampled over time within an individual and analyze their ability to minimize genetic distance to the variants. We show the existence of significant differences in phylogenetic relationship among variants present within an individual and document how it may influence the minimization of genetic distances by different sequences. Based on the number of issues outlined in this study, the use of an autologous viral sequence chosen at random from the pool of variants within an individual is not optimal, and the design of sequences to represent such pools should consider the nature of phylogenetic relationship among the variants.

Materials and methods

HIV-1 infected subject and the env sequences

Pt 8 is a homosexual male enrolled in the Multicenter AIDS Cohort Study (MACS) (Kaslow et al., 1987). Multiple *env* C2V5 sequences, sampled at each of the 13 time points (Visits V09 to V24) over the course of 8 years of infection, with earliest sequences sampled at 4 months following seroconversion (Shankarappa et al., 1999), were used in these analyses. All the sequences used here were sampled prior to the use of highly active combination anti-retroviral therapy in this individual. Pt 8 is among the group of individuals for which we had characterized the virologic (Jensen et al., 2003; Shankarappa et al., 1999; Shriner et al., 2004) and immunologic (Rinaldo et al., 1998) features. Over the course of infection, Pt 8 *env* sequences exhibited strong temporal features. Consistent with the model we had proposed, X4 viruses in this individual evolved at the beginning of stage II, reached a peak, and declined in frequency. Overall, sequences sampled over the course of infection in this subject represented a broad spectrum of viral genetic changes, including a wide range of viral diversity, divergence from early sequences, evolution of X4 viruses, and varying phylogenetic relationships. An average of 12 sequences with open reading frames was available from each visit. Sequences from peripheral blood mononuclear cells (PBMC) were available from all the 13 time points, and sequences from plasma were also available from three of these time points (V11, V15, V19).

Derivation of the consensus, ancestral, and COT sequences

Sequences were aligned and edited to maintain the open reading frame across gaps. Sequences corresponding to each time point were stripped from the same overall alignment and analyzed separately to derive the consensus, ancestral, and COT sequences. For the consensus sequence, we first derived the majority consensus DNA sequence for each time point. To prevent superfluous amino acid resulting from the derivation of consensus nucleotide sequence, each codon in the consensus DNA sequence was checked, compared to the consensus amino acid, and edited.

Ancestral sequences represent the most recent common ancestor (MRCA) for the pool of variants and were derived in PAML using the codon-based maximum likelihood methods (Yang, 1997). Derivation of MRCA for each time point used sequences from the time point along with a set of nine phylogenetically unrelated prototype HIV-1 sequences as the outgroup: HXB2 (K03455), JRFL (U63632), RF (M17451), YU2X (M93258), SF2 (BD016767), NY5 (M38431), CAM1 (D10112), OYI (M26727), MNTQ (AF075719) (GenBank accession numbers indicated in parenthesis). The starting trees for sequences from each time point were derived using maximum likelihood estimates for evolutionary parameters by a two-step process using Paup* (Swofford, 2002), as described by Anderson et al. (2001). Model parameters used in these derivations included the codon frequencies estimated from the average nucleotide frequencies at the three-codon positions. Discrete model for d_N/d_S ratio (ω) was set to vary among sites with three categories for the ω distribution. For the derivation of COT sequences, a maximum likelihood tree was estimated in PAUP*, and the center of the tree, the point in the tree having the least squared distance to the branch tips (Nickle et al., 2003a, 2003b, 2003c), was estimated and the corresponding sequence reconstructed.

Comparing centralized representative sequences to individual variants and identifying positively selected amino acid sites

The dataset of individual viral sequences and the corresponding reconstructed consensus, ancestral, and COT sequences from each time point were aligned and edited at the codon level to derive a master alignment. This realignment maximized the match for insertions and deletions that appeared over the course of infection and was necessary for reconstructing the phylogenetic relationship for the entire dataset. Kimura-2 parameter model of evolution and pairwise gap deletion were used in Mega ver 2.1 (Kumar et al., 2001) to estimate intra-time point diversity (average of all pairwise DNA distances) as well as the DNA distances between individual variants and the consensus, ancestral, and COT sequences from each time point. Similarly, amino acid distances were estimated as *p*-distances using Mega (Kumar et al., 2001). Such simple models were intentionally chosen to minimize the effect of differing evolutionary parameters for sequences sampled over the course of infection and to allow comparisons with earlier and other published results. However, to reconstruct the phylogenetic relationship between representative sequences and the individual viral sequences, neighbor-joining phylograms were derived using maximum likelihood distances estimated in Paup* (Swofford, 2002). Model parameters for the entire set of sequences in the master alignment were used and were as follows: equilibrium nucleotide frequencies: $f_A = 0.3841$, $f_C = 0.1722$, $f_G = 0.2205$, $f_T = 0.2232$; proportion of invariable sites = 0.1804; shape parameter (α) of the Γ distribution reflecting site-to-site rate variability of variable sites: $\alpha = 0.4205$; and the **R** matrix values: $R_{A\to C} = 1.358$, $R_{A\to G} = 3.584$, $R_{A\to T} = 0.801$, $R_{C\to G} = 0.349$, $R_{C\to T} = 3.358$, $R_{G\to T} = 1$.

It would be desirable to identify and partition mutations of potential biologic importance from the large spectrum typically observed in an individual. Within Env, the biologic functions most associated with viral genetic changes are the recognition by immune system and the cell tropism. In view of the documented predominance of positive selection within sites recognized by the immune effector cells (Liu et al., 2004; Ross and Rodrigo, 2002), we sought to identify positively selected sites in the current dataset. We hypothesized that a de novo identification of such positively selected sites would help identify regions targeted for assessing CTL reactivity and limit the number of peptides needed to be synthesized. To test for positive selection, we used model parameters of evolution applicable to sequences from each time point along with a discrete model for selection with a single d_N/d_S ratio and using PAML (Yang, 1997). Positively selected sites for sequences from each time point were then compared to identify patterns of change over the course of infection.

Acknowledgments

We thank David Nickle, Gerald Learn, and Jim Mullins for suggestions. Supported by NIH grants AI041870, AI27757, DK060619, AI035041, AI055794.

References

Addo, M.M., Yu, X.G., Rathod, A., Cohen, D., Eldridge, R.L., Strick, D., Johnston, M.N., Corcoran, C., Wurcel, A.G., Fitzpatrick, C.A., Feeney, M.E., Rodriguez, W.R., Basgoz, N., Draenert, R., Stone, D.R., Brander, C., Goulder, P.J., Rosenberg, E.S., Altfeld, M., Walker, B.D., 2003. Comprehensive epitope analysis of human immunodeficiency virus type 1 (HIV-1)-specific T-cell responses directed against the entire expressed HIV-1 genome demonstrate broadly directed responses, but no correlation to viral load. J. Virol. 77 (3), 2081–2092.

Allen, T.M., O'Connor, D.H., Jing, P., Dzuris, J.L., Mothe, B.R., Vogel, T.U., Dunphy, E., Liebl, M.E., Emerson, C., Wilson, N., Kunstman, K.J., Wang, X., Allison, D.B., Hughes, A.L., Desrosiers, R.C., Altman, J.D., Wolinsky, S.M., Sette, A., Watkins, D.I., 2000. Tat-specific cytotoxic T lymphocytes select for SIV escape variants during resolution of primary viraemia. Nature 407, 386–390.

Allen, T.M., Altfeld, M., Geer, S.C., Kalife, E.T., Moore, C., O'Sullivan, K.M., Desouza, I., Feeney, M.E., Eldridge, R.L., Maier, E.L., Kaufmann, D.E., Lahaie, M.P., Reyor, L., Tanzi, G., Johnston, M.N., Brander, C., Draenert, R., Rockstroh, J.K., Jessen, H., Rosenberg, E.S., Mallal, S.A., Walker, B.D., 2005. Selective escape from CD8+ T-cell responses represents a major driving force of human immunodeficiency virus type 1 (HIV-1) sequence diversity and reveals constraints on HIV-1 evolution. J. Virol. 79 (21), 13239–13249.

Altfeld, M., Addo, M.M., Shankarappa, R., Lee, P.K., Allen, T.M., Yu, X.G., Rathod, A., Harlow, J., O'Sullivan, K., Johnston, M.N., Goulder, P.J., Mullins, J.I., Rosenberg, E.S., Brander, C., Korber, B., Walker, B.D., 2003. Enhanced detection of human immunodeficiency virus type 1-specific T-cell responses to highly variable regions by using peptides based on autologous virus sequences. J. Virol. 77 (13), 7330–7340.

Anderson, J.P., Rodrigo, A.G., Learn, G.H., Wang, Y., Weinstock, H., Kalish, M.L., Robbins, K.E., Hood, L., Mullins, J.I., 2001. Substitution model of sequence evolution for the human immunodeficiency virus type 1 subtype B gp120 gene over the C2-V5 region. J. Mol. Evol. 53 (1), 55–62.

Cornelissen, M., Mulder-Kampinga, G., Veenstra, J., Zorgdrager, F., Kuiken, C., Hartman, S., Dekker, J., van der Hoek, L., Sol, C., Coutinho, R., Goudsmit, J., 1995. Syncytium-inducing (SI) phenotype suppression at seroconversion after intramuscular inoculation of a non-syncytium-inducing/SI phenotypically mixed human immunodeficiency virus population. J. Virol. 69, 1810–1818.

Doria-Rose, N.A., Learn, G.H., Rodrigo, A.G., Nickle, D.C., Li, F., Mahalanabis, M., Hensel, M.T., McLaughlin, S., Edmonson, P.F., Montefiori, D., Barnett, S.W., Haigwood, N.L., Mullins, J.I., 2005. Human immunodeficiency virus type 1 subtype B ancestral envelope protein is functional and elicits neutralizing antibodies in rabbits similar to those elicited by a circulating subtype B envelope. J. Virol. 79 (17), 11214–11224.

Eigen, M., 1993. Viral quasispecies. Sci. Am. 269 (1), 42–49.

Frahm, N., Korber, B.T., Adams, C.M., Szinger, J.J., Draenert, R., Addo, M.M., Feeney, M.E., Yusim, K., Sango, K., Brown, N.V., SenGupta, D., Piechocka-Trocha, A., Simonis, T., Marincola, F.M., Wurcel, A.G., Stone, D.R., Russell, C.J., Adolf, P., Cohen, D., Roach, T., StJohn, A., Khatri, A., Davis, K., Mullins, J., Goulder, P.J., Walker, B.D., Brander, C., 2004. Consistent cytotoxic-T-lymphocyte targeting of immunodominant regions in human immunodeficiency virus across multiple ethnicities. J. Virol. 78 (5), 2187–2200.

Gao, F., Bhattacharya, T., Gaschen, B., Taylor, J., Moore, J.P., Novitsky, V., Yusim, K., Lang, D., Foley, B., Beddows, S., Alam, M., Haynes, B., Hahn, B.H., Korber, B., 2003. Consensus and ancestral state HIV vaccines-Reply to Nickle et al. Science 299 (5612), 1515–1518.

Gao, F., Korber, B.T., Weaver, E., Liao, H.X., Hahn, B.H., Haynes, B.F., 2004. Centralized immunogens as a vaccine strategy to overcome HIV-1 diversity. Expert Rev. Vaccines 3 (Suppl. 4), S161–S168.

Gao, F., Weaver, E.A., Lu, Z., Li, Y., Liao, H.X., Ma, B., Alam, S.M., Scearce, R.M., Sutherland, L.L., Yu, J.S., Decker, J.M., Shaw, G.M., Montefiori, D.C., Korber, B.T., Hahn, B.H., Haynes, B.F., 2005. Antigenicity and immunogenicity of a synthetic human immunodeficiency virus type 1 group m consensus envelope glycoprotein. J. Virol. 79 (2), 1154–1163.

Gaschen, B., Taylor, J., Yusim, K., Foley, B., Gao, F., Lang, D., Novitsky, V., Haynes, B., Hahn, B.H., Bhattacharya, T., Korber, B., 2002. Diversity considerations in HIV-1 vaccine selection. Science 296 (5577), 2354–2360.

Goulder, P.J., Altfeld, M.A., Rosenberg, E.S., Nguyen, T., Tang, Y., Eldridge, R.L., Addo, M.M., He, S., Mukherjee, J.S., Phillips, M.N., Bunce, M., Kalams, S.A., Sekaly, R.P., Walker, B.D., Brander, C., 2001. Substantial differences in specificity of HIV-specific cytotoxic T cells in acute and chronic HIV infection. J. Exp. Med. 193 (2), 181–194.

Herbeck, J.T., Nickle, D.C., Learn, G.H., Gottlieb, G.S., Curlin, M.E., Heath, L., Mullins, J.I., 2006. Human immunodeficiency virus type 1 env evolves toward ancestral states upon transmission to a new host. J. Virol. 80 (4), 1637–1644.

Ho, D.D., Neumann, A.U., Perelson, A.S., Chen, W., Leonard, J.M., Markowitz, M., 1995. Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. Nature 373, 123–126.

Huang, X.L., Fan, Z., Zheng, L., Borowski, L., Li, H., Thomas, E.K., Hildebrand, W.H., Zhao, X.Q., Rinaldo, C.R., 2003. Priming of human immunodeficiency virus type 1 (HIV-1)-specific CD8+ T cell responses by dendritic cells loaded with HIV-1 proteins. J. Infect. Dis. 187 (2), 315–319.

Jensen, M.A., Li, F.S., van 't Wout, A.B., Nickle, D.C., Shriner, D., He, H.X., McLaughlin, S., Shankarappa, R., Margolick, J.B., Mullins, J.I., 2003. Improved coreceptor usage prediction and genotypic monitoring of R5-to-X4 transition by motif analysis of human immunodeficiency virus type 1 env V3 loop sequences. J. Virol. 77 (24), 13376–13388.

Kaslow, R.A., Ostrow, D.G., Detels, R., Phair, J.P., Polk, B.F., Rinaldo Jr., C.R., 1987. The Multicenter AIDS Cohort Study: rationale, organization, and selected characteristics of the participants. Am. J. Epidemiol. 126 (2), 310–318.

Kaufmann, D.E., Bailey, P.M., Sidney, J., Wagner, B., Norris, P.J., Johnston, M.N., Cosimi, L.A., Addo, M.M., Lichterfeld, M., Altfeld, M., Frahm, N., Brander, C., Sette, A., Walker, B.D., Rosenberg, E.S., 2004. Comprehensive analysis of human immunodeficiency virus type 1-specific CD4 responses reveals marked immunodominance of gag and nef and the presence of broadly recognized peptides. J. Virol. 78 (9), 4463–4477.

Kelleher, A.D., Long, C., Holmes, E.C., Allen, R.L., Wilson, J., Conlon, C., Workman, C., Shaunak, S., Olson, K., Goulder, P., Brander, C., Ogg, G., Sullivan, J.S., Dyer, W., Jones, I.I., McMichael, A.J., Rowland-Jones, S., Phillips, R.E., 2001. Clustered mutations in HIV-1 gag are consistently required for escape from HLA-B27-restricted cytotoxic T lymphocyte responses. J. Exp. Med. 193 (3), 375–386.

Korber, B.T., Brander, C., Haynes, B.F., Koup, R.A., Moore, J.P., Walker, B.D., Watkins, D.I. (Eds.), 2003. HIV Immunology and HIV/SIV Vaccine Database. Los Alamos National Laboratory, Theoretical Biology and Biophysics, Los Alamos, New Mexico, Los Alamos, NM 87545. LA-UR 04-8162.

Korber, B., Frahm, N., Yusim, K., Fischer, W., Walker, B., Brander, C., 2005. T-cell Vaccine and Peptide Reagent Design Based on Restricted Patterns of Variation in Conserved HIV Proteins. AIDS Vaccine 2005. Abstract #61, Montreal, Canda.

Kumar, S., Tamura, K., Jakobsen, I.K., Nei, M., 2001. MEGA2: molecular Evolutionary genetics analysis software for microcomputers. Bioinformatics 17 (12), 1244–1245.

Learn, G.H., Mullins, J.I., 2000. The use of an inferred epidemic ancestral sequence as a vaccine immunogen. 7the Annual International Discussion Meeting on HIV Dynamics and Evolution, Seattle, WA.

Learn, G., Mullins, J.I., 2003. The microbial forensic use of HIV sequences. In: Leitner, T., Foley, B., Hahn, B., Marx, P., McCutchan, F., Mellors, J., Wolinsky, S., Korber, B. (Eds.), Theoretical Biology and Biophysics Group. HIV Sequence Compendium 2003. Los Alamos National Laboratory. LA-UR number 04-7420.

Learn, G.H., Muthui, D., Brodie, S.J., Zhu, T., Diem, K., Mullins, J.I., Corey, L., 2002. Virus population Homogenization following acute human immunodeficiency virus type 1 infection. J. Virol. 76 (23), 11953–11959.

Liu, Y., Zhao, H., Genowati, I., McNevin, J., Nickle, D., Shriner, D., Wong, K., Cao, J., Davies, K., Rose, L., McElrath, M.J., Mullins, J.I., 2004. CTL Responses as Major Selective Forces Shaping the Course of HIV-1 Evolution in vivo. 11th Conference on Retroviruses and Opportunistic Infections, San Francisco, CA.

Mansky, L.M., Temin, H.M., 1995. Lower in vivo mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase. J. Virol. 69 (8), 5087–5094.

Mullins, J.I., Nickle, D.C., Heath, L., Rodrigo, A.G., Learn, G.H., 2004. Immunogen sequence: the fourth tier of AIDS vaccine design. Expert Rev. Vaccines 3 (Suppl. 1), S151–S159.

Myers, G., Korber, B., Wain-Hobson, S., Smith, R.F., Pavlakis, G.N., 1993. Human retroviruses and AIDS, 1993. A Compilation and Analysis of Nucleic Acid and Amino Acid Sequences. Los Alamos National Laboratory.

Nickle, D.C., Jensen, M.A., Gottlieb, G.S., Shriner, D., Learn, G.H., Rodrigo, A.G., Mullins, J.I., 2003a. Consensus and ancestral state HIV vaccines. Science 299 (5612), 1515–1518 (author 2003a).

Nickle, D.C., Jensen, M.A., Shriner, D., Brodie, S.J., Frenkel, L.M., Mittler, J.E., Mullins, J.I., 2003b. Evolutionary indicators of human immunodeficiency virus type 1 reservoirs and compartments. J. Virol. 77 (9), 5540–5546.

Nickle, D.C., Shriner, D., Mittler, J.E., Frenkel, L.M., Mullins, J.I., 2003c. Importance and detection of virus reservoirs and compartments of HIV infection. Curr. Opin. Microbiol. 6 (4), 410–416.

Nielsen, R., Yang, Z., 1998. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics 148 (3), 929–936.

Novitsky, V., Smith, U.R., Gilbert, P., McLane, M.F., Chigwedere, P., Williamson, C., Ndung'u, T., Klein, I., Chang, S.Y., Peter, T., Thior, I., Foley, B.T., Gaolekwe, S., Rybak, N., Gaseitsiwe, S., Vannberg, F., Marlink, R., Lee, T.H., Essex, M., 2002. Human immunodeficiency virus type 1 subtype C molecular phylogeny: consensus sequence for an AIDS vaccine design? J. Virol. 76 (11), 5435–5451.

O'Connor, D.H., Allen, T.M., Vogel, T.U., Jing, P., DeSouza, I.P., Dodds, E., Dunphy, E.J., Melsaether, C., Mothe, B., Yamamoto, H., Horton, H., Wilson, N., Hughes, A.L., Watkins, D.I., 2002. Acute phase cytotoxic T lymphocyte escape is a hallmark of Simian immunodeficiency virus infection. Nat. Med. 8 (5), 493–499.

Pathak, V.K., Temin, H.M., 1990a. Broad spectrum of in vivo forward mutations, hypermutations, and mutational hotspots in a retroviral shuttle vector after a single replication cycle: deletions and deletions with insertions. Proc. Natl. Acad. Sci. U.S.A. 87 (16), 6024–6028.

Pathak, V.K., Temin, H.M., 1990b. Broad spectrum of in vivo forward mutations, hypermutations, and mutational hotspots in a retroviral shuttle vector after a single replication cycle: substitutions, frameshifts, and hypermutations. Proc. Natl. Acad. Sci. U.S.A. 87 (16), 6019–6023.

Poss, M., Rodrigo, A.G., Gosink, J.J., Learn, G.H., de Vange Panteleeff, D., Martin Jr., H.L., Bwayo, J., Kreiss, J.K., Overbaugh, J., 1998. Evolution of envelope sequences from the genital tract and peripheral blood of women infected with clade A human immunodeficiency virus type 1. J. Virol. 72 (10), 8240–8251.

Rinaldo Jr., C.R., Gupta, P., Huang, X., Fan, Z., Mullins, J.I., Gange, S., Farzadegan, H., Shankarappa, R., Muñoz, A., Margolick, J.B., 1998. Anti-HIV-1 memory cytotoxic T lymphocyte responses associated with changes in CD4+ T cell numbers in the progression of HIV-1 infection. AIDS Res. Hum. Retrovir. 14 (16), 1423–1433.

Roberts, J.D., Bebenek, K., Kunkel, T.A., 1988. The accuracy of reverse transcriptase from HIV-1. Science 242 (4882), 1171–1173.

Ross, H.A., Rodrigo, A.G., 2002. Immune-mediated positive selection drives human immunodeficiency virus type 1 molecular variation and predicts disease duration. J. Virol. 76 (22), 11715–11720.

Shankarappa, R., Margolick, J.B., Gange, S.J., Rodrigo, A.G., Upchurch, D., Farzadegan, H., Gupta, P., Rinaldo, C.R., Learn, G.H., He, X., Huang, X.L., Mullins, J.I., 1999. Consistent viral evolutionary changes associated with the progression of human immunodeficiency virus type 1 infection. J. Virol. 73 (12), 10489–10502.

Shriner, D., Shankarappa, R., Jensen, M.A., Nickle, D.C., Mittler, J.E., Margolick, J.B., Mullins, J.I., 2004. Influence of random genetic drift on human immunodeficiency virus type 1 env evolution during chronic infection. Genetics 166 (3), 1155–1164.

Swofford, D.L., 2002. PAUP* 4.0: Phylogenetic Analysis Using Parsimony (And other methods). ver 4.0b10. 2002.

van 't Wout, A.B., Kootstra, N.A., Mulder-Kampinga, G.A., Albrecht-van Lent, N., Scherpbier, H.J., Veenstra, J., Boer, K., Coutinho, R.A., Miedema, F., Schuitemaker, H., 1994. Macrophage-tropic variants initiate HIV-1 infection after sexual, parenteral, and vertical transmission. J. Clin. Invest. 94 (5), 2060–2067.

Wei, X., Ghosh, S.K., Taylor, M.E., Johnson, V.A., Emini, E.A., Deutsch, P., Lifson, J.D., Bonhoeffer, S., Nowak, M.A., Hahn, B.H., Saag, M.S., Shaw, G.M., 1995. Viral dynamics of HIV-1 infection. Nature 373 (6510), 117–122.

Williamson, S., 2003. Adaptation in the env gene of HIV-1 and evolutionary theories of disease progression. Mol. Biol. Evol. 20 (8), 1318–1325.

Wolfs, T.F., Zwart, G., Bakker, M., Goudsmit, J., 1992. HIV-1 genomic RNA diversification following sexual and parenteral virus transmission. Virology 189, 103–110.

Wolinsky, S.M., Wike, C.M., Korber, B.T.M., Hutto, C., Parks, W.P., Rosenblum, L.L., Kunstman, K.J., Furtado, M.R., Muñoz, J.L., 1992. Selective transmission of HIV-1 variants from mother to infants. Science 255, 1134–1137.

Wolinsky, S.M., Korber, B.T.M., Neumann, A.U., Daniels, M., Kuntsman, K.J., Whetsell, A.J., Furtado, M.R., Cao, Y., Ho, D.D., Safrit, J.T., Koup, R.A., 1996. Adaptive evolution of HIV-1 during the natural course of infection. Science 272, 537–542.

Yamaguchi-Kabata, Y., Gojobori, T., 2000. Reevaluation of amino acid variability of the human immunodeficiency virus type 1 gp120 envelope glycoprotein and prediction of new discontinuous epitopes. J. Virol. 74 (9), 4335–4350.

Yang, Z., 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput. Appl. Biosci. 13 (5), 555–556.

Yu, X.G., Addo, M.M., Rosenberg, E.S., Rodriguez, W.R., Lee, P.K., Fitzpatrick, C.A., Johnston, M.N., Strick, D., Goulder, P.J., Walker, B.D., Altfeld, M., 2002. Consistent patterns in the development and immunodominance of human immunodeficiency virus type 1 (HIV-1)-specific CD8+ T-cell responses following acute HIV-1 infection. J. Virol. 76 (17), 8690–8701.

Yusim, K., Kesmir, C., Gaschen, B., Addo, M.M., Altfeld, M., Brunak, S., Chigaev, A., Detours, V., Korber, B.T., 2002. Clustering patterns of cytotoxic T-lymphocyte epitopes in human immunodeficiency virus type 1 (HIV-1) proteins reveal imprints of immune evasion on HIV-1 global variation. J. Virol. 76 (17), 8757–8768.

Zanotto, P.M., Kallas, E.G., de Souza, R.F., Holmes, E.C., 1999. Genealogical evidence for positive selection in the nef gene of HIV-1. Genetics 153 (3), 1077–1089.

Zhang, L.Q., MacKenzie, P., Cleland, A., Holmes, E.C., Leigh Brown, A.J., Simmonds, P., 1993. Selection for specific sequences in the external envelope protein of HIV-1 upon primary infection. J. Virol. 67 (6), 3345–3356.

Zhao, X.Q., Huang, X.L., Gupta, P., Borowski, L., Fan, Z., Watkins, S.C., Thomas, E.K., Rinaldo, C.R., 2002. Induction of anti-human immunodeficiency virus type 1 (HIV-1) CD8(+) and CD4(+) T-cell reactivity by dendritic cells loaded with HIV-1 X4-infected apoptotic cells. J. Virol. 76 (6), 3007–3014.

Zhu, T., Mo, H., Wang, N., Nam, D.S., Cao, Y., Koup, R.A., Ho, D.D., 1993. Genotypic and phenotypic characterization of HIV-1 in patients with primary infection. Science 261, 1179–1181.

Zhu, T., Wang, N., Carr, A., Nam, D.S., Moor-Jankowski, R., Cooper, D.A., Ho, D.D., 1996. Genetic characterization of human immunodeficiency virus type 1 in blood and genital secretions: evidence for viral compartmentalization and selection during sexual transmission. J. Virol. 70 (5), 3098–3107.